Recent Innovations in the Changing Criterion Design:

Implications for Research and Practice in Special Education

Dennis McDougall, *University of Hawai'i*
Jacqueline Hawkins, *University of Houston*
Michael Brady, *Florida Atlantic University*
Amelia Jenkins, *University of Hawai'i*

This article illustrates (a) 2 recent innovations in the changing criterion research design, (b) how these innovations apply to research and practice in special education, and (c) how clinical needs influence design features of the changing criterion design. The first innovation, the range-bound changing criterion, is a very simple variation of the classic changing criterion design. The classic version uses a single criterion for each stepwise intervention phase, whereas the range-bound version uses a range criterion—that is, an upper and lower limit for each intervention phase. The second innovation, the distributed criterion, combines elements of the changing criterion, multiple-baseline, and ABAB designs. It is well suited to contexts where students must multitask—that is, allocate, prioritize, and adjust time and effort to complete multiple tasks in response to changing environmental demands. These two innovations expand options available to researchers who use single-case research designs to investigate questions of interest in special education.

In 1982, Schloss, Sedlack, Elliott, and Smothers published an article in *The Journal of Special Education* that illustrated "applications of the changing criterion designs to special education classrooms" (p. 361). These applications were based on the classic changing criterion (CC) research design, which Hartmann and Hall (1976) said was "initially named by Hall (1971) and illustrated by Weis and Hall (1971) [and] described, but unnamed by Sidman (1960, pp. 254–256)" (p. 527). For nearly half of a century, researchers have used the CC and other classic single-case designs, particularly the ABAB and multiple-baseline research designs, in special education and other settings. These single-case designs are very useful for evaluating experimental control in studies that (a) include one or a few students; (b) require ongoing, repeated, and quantitative measures of individual students' progress across time; and (c) apply interventions that seek to improve students' performance of socially valid, directly observable, and measurable target behaviors.

Development of Single-Case Designs and Applied Behavior Analysis

In the 1960s, while investigators were developing classic single-case research designs, applied behavior analysis was emerging as a behavior-change technology and as a methodology for evaluating experimental control of interventions that promote intraindividual change over time (Baer, Wolf, & Risley, 1968). Numerous innovations of these classic designs subsequently appeared. According to Hartmann and Hall (1976), "The development of experimental designs to demonstrate control in individual case studies has been a crucial factor in bringing about scientific status to the study of individuals" (p. 527). In many cases, researchers designed these innovations to accommodate their research questions, conditions of the specific intervention and target behavior, and ethical and clinical considerations, or to demonstrate experimental control via visual inspection of graphed data (McLaughlin, 1983).

Kazdin (1982) described numerous variations on the classic single-case research designs, as well as the treatment evaluation strategies and outcome questions that corresponded to these variations. Kazdin concluded, however, that "few variations of the changing criterion design have been developed" (p. 159). During the two decades since Kazdin reached this conclusion, few variations of the CC design have been created (McDougall, 2005b; McDougall, Smith, Black, & Rumrill, 2005). Recently, McDougall (2005a, 2005c) developed and applied two innovations of the classic CC design, which he called the *range-bound changing criterion* (RBCC) and the *distributed criterion* (DC). Our purpose in this article is to introduce these two design innovations to the field of special education. We describe how investigators applied these design innovations in the two studies that have used them as of this writing. We show how researchers and practitioners can apply these inno-

Address: Dennis McDougall, University of Hawai'i, Department of Special Education, 124 Wist Hall, 1776 University Ave., Honolulu, HI 96822; e-mail: mcdougal@hawaii.edu

vations and associated strategies to special education and illustrate how clinical needs influence design features. First, however, we review the classic CC design and how investigators have used this design to answer questions of interest in special education and related disciplines.

The Classic Changing Criterion Research Design

Although used less frequently than multiple-baseline and ABAB designs, the classic CC design has been applied by researchers in special education and related disciplines to numerous single-case studies that targeted behaviors of children and teenagers with and without disabilities. Examples include the following uses:

1. to increase the amount of time a boy with separation anxiety disorder was able to be away from his mother (Flood & Wilder, 2004);
2. to increase a teenager's compliance with a medical regimen (Gorski & Westbrook, 2002);
3. to increase the number of math problems completed correctly by children (Hall & Fox, 1977; Schloss, Sedlak, Elliott, & Smothers, 1982); and
4. to increase the food acceptance of individuals with chronic food refusal (Kahng, Boscoe, & Byrne, 2003; Luiselli, 2000).

Researchers also have targeted behaviors of adults. Examples include the following:

1. to increase the amount of leisure-time reading of an adult diagnosed with schizophrenia (Skinner, Skinner, & Armstrong (2000);
2. to increase the work rate of adults with severe/profound mental retardation (Bates, Renzaglia, & Clees, 1980);
3. to reduce adults' cigarette smoking (Edinger, 1978; Weis & Hall, 1971); and
4. to reduce an adult's excessive coffee drinking (Foxx & Rubinoff, 1979).

In addition, some researchers have embedded or combined the CC design with the multiple-baseline design (cf. Hinerman, Jenson, Walker, & Petersen, 1982; Mizes, 1985; Noles, Epstein, & Jones, 1976; Paniagua, Pumariega, & Black, 1988; Schleien, Wehman, & Kiernan, 1981).

The classic CC design is most appropriate for evaluating the effects of interventions that aim to change—in a therapeutic direction (i.e., accelerate or decelerate)—one target behavior of one research participant in a systematic, stepwise fashion. The CC design contributed greatly to the field because it overcame access-to-treatment issues inherent in two existing single-case designs. That is, the CC provided another option, besides the multiple baseline design, for demonstrating experimental control without requiring reversals or withdrawing treatment, as in the ABAB. The CC design also did not require researchers to delay treatment, as is the case for the multiple-baseline design, which staggers or lags treatment. Researchers and practitioners in special education have found the CC design useful as an instructional approach and a research design when the target behavior and corresponding intervention lend themselves to setting explicit standards or performance criteria. Experimental control for CC designs is demonstrated when changes in the target behavior match precisely, or correspond closely to, at least three shifts in performance criteria. However, experimental control is more difficult to evaluate when the target behavior substantially exceeds performance criteria.

Researchers who use the RBCC and DC designs should adhere to experimental control and design guidelines that apply to CC designs. First, they must establish performance criteria a priori, that is, before they institute each successive, stepwise intervention phase. Setting a priori, rule-governed performance criteria is consistent with scientific goals that require researchers to describe, explain, predict, and control phenomena and to replicate intervention impact (Kerlinger, 1986). Second, researchers should judiciously shift performance criteria across adjacent phases so that (a) the *magnitude* of behavior change across adjacent phases is large enough to demonstrate experimental control, (b) changes in the *level* of performance conform closely to changes in performance criteria, and (c) *latency* of change is minimal. Third, researchers must collect enough data within each phase to enable conclusions about the stability of behavior changes. The target behavior should show minimal variability and a flat or countertherapeutic trend before a criterion change is instituted. Fourth, researchers must shift performance criteria and replicate behavior change at least three times. Finally, researchers should revert and then reinstitute performance criteria, *if necessary and appropriate,* to bolster experimental control when the target behavior deviates from performance criteria.

The Range-Bound Changing Criterion

The RBCC is a very simple variation of the classic CC design, with one differentiating feature. Within respective stepwise intervention phases, the CC uses a *single* performance criterion, whereas the RBCC uses *both a lower criterion and an upper criterion*. In the RBCC, the target behavior must match or exceed the lower performance criterion *and*, *concurrently*, match or be less than the higher performance criterion. The two criteria define a range of expected performance. This contrasts with the CC, in which a participant's behavior is expected to (a) match or exceed the single-point criterion

for an intervention phase when the intervention aims to increase a behavior or (b) match or remain lower than the criterion for an intervention phase when the intervention aims to decrease a behavior. In the CC, success is achieved when actual performance matches or supersedes the stipulated, single-point criterion (e.g., play at recess with three or more peers). Conversely, in the RBCC, success is achieved when a participant's behavior resides within, or conforms to, the stipulated range criterion (e.g., play at recess with at least three, but not more than four, peers).

Both the CC and RBCC typically include an initial baseline phase followed by a series of intervention phases, each of which has a stepwise (i.e., changing) criterion for performance and serves as a baseline for subsequent phases of the intervention. Both versions are applicable for interventions in which the aim is to accelerate one target behavior (e.g., daily exercise) or decelerate one target behavior (e.g., cursing) in one context, typically in a criterion-determined, systematic, and sequential fashion. Thus, both versions lend themselves to procedures such as (a) shaping and differential reinforcement of higher or lower rates of behavior (Alberto & Troutman, 1999); (b) cognitive-behavioral modification, including goal setting with frequent feedback (Kottler, 2001); and (c) behavioral self-management procedures, such as self-monitoring, self-verbalization, self-evaluation, and self-graphing (Glynn, Thomas, & Shee, 1973; McDougall, 1998; Watson & Tharp, 2002).

First Application of the RBCC

In the first study to apply the RBCC (McDougall, 2005a), an overweight adult used goal setting and behavioral self-management to increase the duration of daily exercise, reduce body weight, and improve cardiovascular functioning. The primary target behavior, daily exercise, was operationally defined and measured as the number of minutes (duration) that the participant ran on a daily basis. Goal setting required that the participant establish long-term and intermediate goals (e.g., lose 40 pounds within 1 year), as well as a series of short-term objectives with explicit, stepwise performance criteria. Behavioral self-management required the participant to graph, each day, the number of minutes spent running.

Establishing Expected Performance Ranges. Table 1 and Figure 1 depict how many minutes per day the participant expected to run (via goal setting) and actually ran during the baseline, intervention, and maintenance phases. Prior to each intervention phase, the participant decided how many minutes he would run each day, during a 6-day week (1 day per week was a rest day). For example, during the first intervention phase, the participant aimed to run, on average, 20 min per day for 6 days each week. Moreover, the participant established around this mean a range of ±10% to ascertain the minimum and maximum number of minutes he should run each day. This pre-established range of expected performance (18–22 min/day) for the first intervention phase is indicated by the two solid and parallel horizontal lines in Figure 1. Thus, the participant's first short-term objective was to run at least 18 min—but no more than 22 min—for each of the 6 exercise days of the initial intervention phase. After mastering this first objective, the participant established a second objective and increased the criterion a mean of 40 (±4) min per day for the second intervention phase; then 60 (±6) min per day for the third intervention phase, and so on.

Rationale for Establishing RBCC. Why establish range criteria rather than a single criterion for each intervention phase? In McDougall (2005a), the primary rationale for establishing a range was clinical rather than experimental. The higher of the two within-phase criteria established a ceiling above which the participant aimed *not* to run. This upper limit reduced the likelihood of injuries due to running too much and too soon during a running program designed to increase the participant's cardiovascular endurance gradually. In the past, the participant had injured himself repeatedly when he interspersed very long runs during periods when the duration of typical daily running was relatively short. Second, the lower of the two within-phase criteria established a floor below which the participant aimed *not* to run. This lower limit reduced the likelihood that the participant would run for no or just a few minutes—a duration that would contribute little to the goal of gradually increasing cardiovascular endurance. This lower limit also increased the likelihood of successfully shaping longer runs, which was an outcome that was consistent with the participant's long-term goals of running a marathon and losing body weight. Finally, the combination of lower and higher criteria ensured a relatively consistent range of expected performance while allowing some flexibility or variation from day to day. On days when the participant felt somewhat tired, he could opt to run up to 10% less than the mean criterion for that phase. On days when the participant felt relatively strong, he could opt to run as much as 10% more than the mean criterion for that phase.

Demonstrating Experimental Control. As evidenced by the data provided in Figure 1 and Table 1, the intervention demonstrated strong experimental control over the target behavior from baseline through the first five intervention phases. After running only 3 days during 19 weeks of baseline (M_B = 0.8 min/day), the duration of running shifted in stepwise fashion during the initial intervention phase ($M_{I,\ 18–22min}$ = 20.0 min/day), the next three intervention phases ($M_{I,\ 36–44min}$ = 41.3 min/day, $M_{I,\ 54–66min}$ = 60.9 min/day, and $M_{I,\ 72–88min}$ = 80.8 min/day, respectively), and the subsequent phase, when performance criteria reverted ($M_{Irevert,\ 54–66min}$ = 63.9 min/day). During the next phase ($M_{Irevert,\ 72–88min}$ = 80.4), the intervention continued to demonstrate functional control, even though an ascending trend appeared in the last half of the preceding phase, because the data during this sixth

TABLE 1. Changes in Running Duration Conform to Changes in Performance Criteria During Intervention

Phase	Within-phase criteria Range	Within-phase criteria *M*	Actual performance Range	Actual performance *M*	PCD
Baseline	*na*	*na*	0.0–40.0	0.8	*na*
Intervention 1	18–22/day	20.0	0.0	20.0	6/6 = 100%
Intervention 2	36–44/day	40.0	37.5–44.0	41.3	6/6 = 100%
Intervention 3	54–66/day	60.0	57.3–66.0	60.9	6/6 = 100%
Intervention 4	72–88/day	80.0	72.5–88.0	80.8	6/6 = 100%
Intervention 5	54–66/day	60.0	57.0–66.0	63.9	8/8 = 100%
Intervention 6	72–88/day	80.0	73.2–88.0	80.4	6/6 = 100%
Intervention 7	80–120/day	100.0	83.2–120.0	103.5	6/6 = 100%
Maintenance	*na*	*na*	13.0–140.0	90.5	*na*

Note. All values are running duration in minutes, except data reported in PCD column. PCD (percentage of conforming data) = the percentage of data points that conformed to a priori, within-phase, range criteria for respective intervention phases. *na* = not applicable. From "The Range-Bound Changing Criterion Design," by D. McDougall, 2005, *Behavioral Interventions, 20,* p. 132. Copyright 2005 by Wiley. Reprinted with permission.

intervention phase were stable and conformed to the range criteria. Running duration continued to increase during the final intervention phase ($M_{I,\ 80\text{–}120min}$ = 103.5 min/day), although data were variable and one data point overlapped with the preceding phase.

Percentage of Conforming Data. When using the RBCC, researchers should analyze the extent to which data points within intervention phases conform to a priori criteria that establish or "bound" the range of acceptable performance for the respective intervention phases. In McDougall (2005a), this analysis further verified the extent of experimental control; that is, the number of minutes the participant actually ran conformed very closely to the within-phase ranges established a priori for each intervention phase. As depicted in Figure 1 and Table 1, every data point ($n = 44$) within each of the seven intervention phases conformed to within-phase criteria. That is, 6 of 6 data points within the initial intervention phase resided within the predetermined range of 18 to 22 min, 6 of 6 data points resided within the predetermined range of 36 to 44 min for the next phase, and so forth, through the last intervention phase, when 6 of 6 data points resided within the predetermined range of 80 to 120 min.

We recommend that researchers use and report an index that quantifies the extent to which data points conform to range criteria. The index, percentage of conforming data (PCD), is calculated using the following formula: PCD = number of data points within intervention phases that conform (i.e., reside within a priori criterion ranges for respective intervention phases), divided by total number of data points within all range-bound intervention phases, and multiplied by 100%. Again, as seen in Table 1 and Figure 1, 44 of 44 data points from the seven intervention phases in the McDougall (2005a) study conformed to their respective a priori ranges; thus, PCD equaled 100%. The PCD index becomes weaker when data points fail to conform to (i.e., reside above or below) criterion ranges. Finally, investigators should use the PCD index to supplement, rather than replace, visual inspection criteria (Kazdin, 1982) as a tool for evaluating experimental control.

Applying the RBCC in Special Education Research and Practice

General Case. We can envision many research questions, interventions, and target behaviors in special education in which researchers might apply the RBCC instead of the CC. First, consider situations in which sudden and excessive fluctuations in the target behavior during intervention phases—even when in the desirable, therapeutic direction—might inhibit short-term performance, long-term change, or maintenance. In such cases, researchers, practitioners, and participants might wish to ensure that the target behavior occurs consistently within a reasonable range (in other words, without excessive variability). Second, consider interventions designed to permit or promote "healthy variability" in the target behavior. Such variability would be associated with better outcomes when compared to situations where excessive variability might occur, such as when we specify a single-point criterion, which does not restrict, for example, the upper limit of the participants' performance. In other contexts, reasonable variability might be better than no variability where the target behavior never deviates from a single-point criterion.

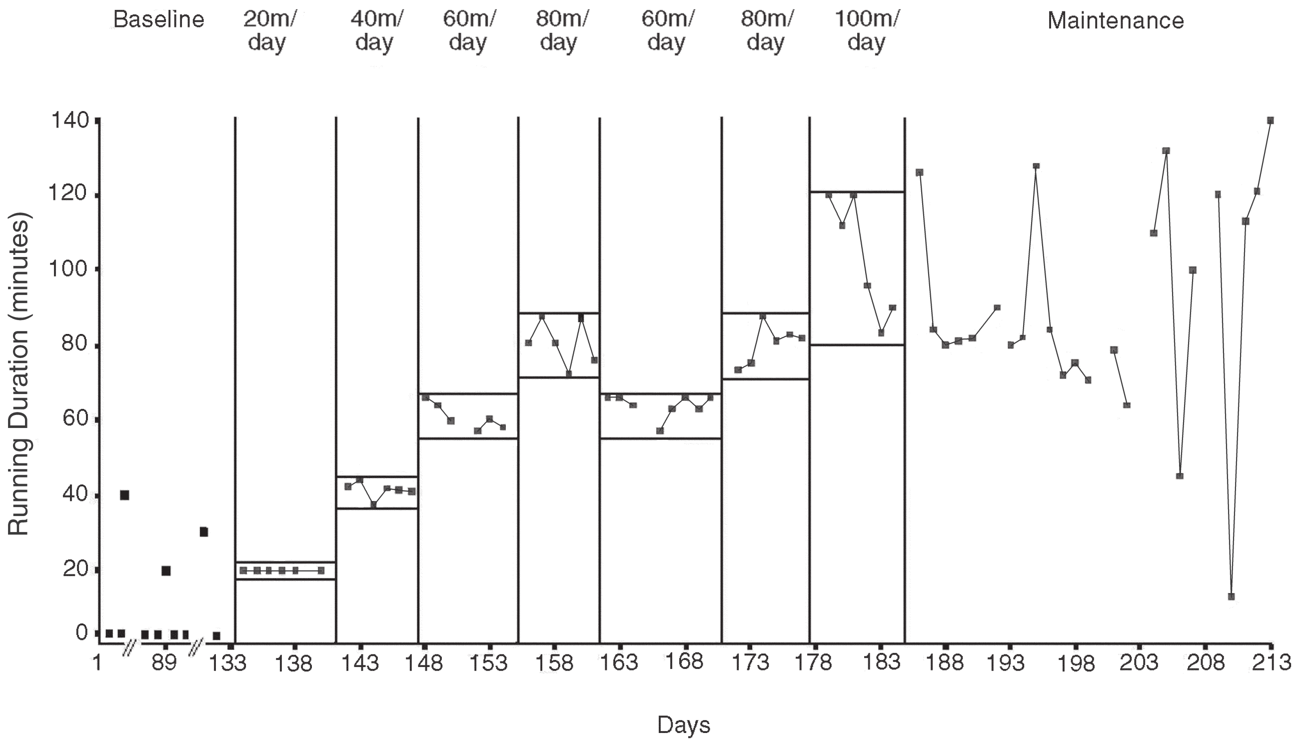

FIGURE 1. Duration, in minutes, of daily exercise (i.e., running) during the baseline, intervention, and maintenance phases. *Note.* Baseline is truncated to permit display of all data from intervention and maintenance phases. Participant ran on 3 days during 19-week baseline phase. Parallel horizontal lines within the seven intervention phases depict performance criteria that participant established for self to define the range of acceptable performance for each phase. Upper horizontal line indicates the maximum number of minutes of running permitted, lower horizontal line indicates the minimum number of minutes of running permitted. From "The Range-Bound Changing Criterion Design," by D. McDougall, 2005, *Behavioral Interventions, 20,* p. 132. Copyright 2005 by Wiley. Reprinted with permission.

Specific Examples. Possible applications of the RBCC in special education research include interventions designed to improve, in a stepwise fashion, social greetings initiated by isolated or withdrawn students; daily exercise by students with mobility impairments, weight problems, or sedentary lifestyles; academic productivity of students who complete tasks too slowly; and academic accuracy of students who rush through their independent assignments with high error rates. We will look in more detail at interventions that aim to increase the number of greetings that a socially withdrawn student initiates while in the presence of peers in large-group settings, such as the cafeteria, playground, school-wide assemblies, hallways, and the school bus. Interventions such as contingent reinforcement, differential reinforcement of higher rates of behavior, goal setting, and self-monitoring, when applied via a CC approach, could increase the frequency of the student's greetings. However, excessive greetings, extremely variable frequencies of greetings from day to day, or an occasional spike in the number of greetings might produce undesirable, unintended outcomes. Indeed, other students might interpret these fluctuations as weird. In this case, it would be wise to specify a reasonable *range* for frequency of greetings within intervention phases and to move stepwise from no or few occurrences toward a terminal goal. The RBCC could help to prevent excessive variability and undesirable outcomes and ensure that greetings shifted systematically and remained consistently within a socially acceptable and personally rewarding range.

The RBCC approach also might be advantageous for students who are prone to noncompliance, for example, students with behavior disorders who react poorly to adult directives that are in the form of an absolute, single-point criterion (e.g., do 12 math problems, clean the tabletops for 10 min). Some students might respond more favorably to adult demands stated in the form of a range (e.g., do 10–14 math problems; clean the tabletops for at least 8, but no more than 12, min). The RBCC approach meshes well with directives and activities that permit teachers and students some flexibility or choice. The RBCC also might mediate against always just meeting minimum standards.

Establishing Effective Ranges

What is an effective range to specify as the criterion for expected performance of the target behavior? How wide or narrow should ranges be for the terminal criterion and for the intermediate criteria that precede the terminal criterion? The answers depend, in part, on who asks these questions—a researcher who seeks to demonstrate experimental control or a practitioner who aims to achieve clinical improvements in student behavior. In some cases, researchers and practitioners are likely to agree when establishing effective ranges; in other cases, competing interests might make it difficult to establish ranges that are acceptable to each party.

Social Validity. One approach for establishing effective ranges for performance criteria, particularly terminal goals, is to apply the logic inherent in social validity methods, including the social comparison and subjective evaluation methods (Kazdin, 1982). With social comparison, investigators could use direct observation to ascertain the range of the target behavior that participants' peers exhibit under conditions similar to those in which participants are expected to perform. These data would reveal, for example, the typical or normative range of greetings for third-grade girls who play during recess at a particular school: Do these girls most typically tend to greet one to three peers on the playground, three to five peers, or five to seven peers?

Alternatively, investigators could use the subjective evaluation method to establish effective ranges, particularly for terminal goals. They could ask recess supervisors, "What's the typical range of greetings for most third-grade girls during recess?" and also ask multiple third-grade girls (i.e., the participants' peers), "How many friends do you usually greet or say hello to during recess?" The investigators could formulate a reasonable range for the targeted student based on the distribution of responses from peers (e.g., five third-grade girls who stated that they usually greeted about two to three, four to six, four to five, five to six, or six to nine of their classmates). Investigators also could ask targeted students what they thought the range should be, particularly in studies that include goal setting or self-management interventions. Seeking input directly from participants is conducive to principles of self-determination. In any event, the essential task is to establish a range that permits *enough but not too much variability*. Some target behaviors might be amenable to narrow ranges, whereas others might be conducive to wide ranges.

A range-bound criterion can provide additional opportunities to analyze performance. Teachers can use results, particularly in graphed form, to teach students valuable lessons about behavior change, goal setting, effort, and outcomes, including how other people perceive these students' behavior patterns. One example would be possible patterns of performance from an intervention that successfully increases the number of correct responses a student produces during independent math practice. If we assume that the student's target behavior always resides within the range-bound criteria (PCD = 100%), this would be good news from the standpoint of experimental control, particularly when the within-phase ranges are narrow and do not overlap. However, researchers and practitioners must interpret such results carefully, because variations in performance are possible even when PCD = 100%. For example, how might we interpret and use data when a student's target behavior *always matches the lower limit* of the range criteria, *always matches the upper limit* of the range criteria, or *fluctuates somewhat*, sometimes tending toward the lower limit and sometimes tending toward the upper limit of the range criteria? Teachers could utilize the graphed data to instruct students about the pros and cons of always performing in ways that just meet the minimum acceptable criterion; always performing in ways that meet the maximum recommended criterion; and adjusting or prioritizing their efforts such that performance sometimes meets the minimum acceptable criterion and, at other times, meets the maximum recommended criterion.

Experimental Control. To maximize experimental control, we recommend that researchers attend closely to the following guidelines when they establish ranges for performance criteria. First, establish relatively narrow ranges for each intervention phase. Narrow ranges are consistent with the need to demonstrate *stable* performance within intervention phases. If researchers establish ranges that are too wide, variability of performance that might be acceptable clinically and is consistent with a therapeutic goal could threaten internal validity and weaken experimental control. As seen in Figure 1, in the McDougall (2005a) study, the target behavior always conformed to the narrow-range criterion (±10% of a mean) that had been stipulated for Intervention Phases 1 through 6 (PCD = 100%). Performance thus was consistent and stable. During the seventh (i.e., last) intervention phase, target behavior also always conformed to the stipulated range criterion (PCD = 100%); however, this criterion (±20% of a mean) was wider than in all preceding intervention phases, and performance was quite variable. Combined with patterns of stable performance during Intervention Phases 1 through 6 and variable performance during Phase 7, extensive variability during maintenance—when the participant continued to self-graph his behavior but did not utilize the performance criteria—suggests that these criteria exerted some control over the target behavior, particularly in regards to stability.

Second, researchers should establish range criteria that do not *overlap* in adjacent phases. This will reduce the likelihood of obtaining overlapping data between adjacent intervention phases. In the McDougall (2005a) study, during the first six intervention phases the range criteria for adjacent phases did not overlap, the target behavior always resided within (adhered to) the range criteria established for the respective intervention phases, zero overlap in the target behavior existed between adjacent intervention phases, and experimental control was quite strong (see Table 1 and Fig-

ure 1). Conversely, experimental control was weakened somewhat, as evidenced by 1 data point (83 min) in the seventh intervention phase that overlapped with highest data point (88 min) in the sixth intervention phase. The range criteria for the seventh intervention phase (80–120 min) overlapped with the range criteria for the sixth intervention phase (72–88 min). Note also that the target behavior always conformed to the predetermined range criteria for the sixth and seventh intervention phases (PCD = 100%); however, the overlap between performance criteria for these phases "permitted" the participant to perform the target behavior in a manner that weakened experimental control. In this case, therapeutic goals and experimental control standards conflicted: *overlap* between adjacent intervention phases occurred for the first time during Phase 7, when the researcher widened the permissible range for running.

Summary of the RBCC

The RBCC is a simple variation of the classic CC design. The differentiating feature of the former is the requirement to specify a bounded range of expected performance rather than a single-point criterion. This range is operationally defined by specifying an upper criterion and a lower criterion. Consequently, the RBCC offers one additional way to evaluate experimental control that is not available when using the CC design. That is, the RBCC requires researchers to ask, "To what extent does the target behavior conform to the range specified by the within-phase criteria for minimum and maximum performance?" The case for experimental control is clearest and strongest when the stipulated ranges are narrow and do not overlap across adjacent phases, and the target behavior resides consistently within the range defined by the upper and lower criteria for each intervention phase. As illustrated previously, researchers can calculate the simple PCD index and report the percentage of data points that conform to the respective a priori range criteria. Thus, one potential advantage of the RBCC in evaluating functional control is a comparatively clearer standard for evaluating the stability–variability of graphed data within each intervention phase.

Cautions

As illustrated in Figure 1, the RBCC can accommodate temporary reversals of direction in performance criteria to permit additional opportunities to evaluate functional control. However, as with the CC, when investigators temporarily change performance criteria in the opposite direction from that of the usual stepwise one, this is by definition a nontherapeutic change. Researchers therefore should exercise caution and weigh the ethical, practical, and experimental pros and cons of instituting these temporary reversals of direction in performance criteria. For some target behaviors, temporarily reverting performance criteria to previous levels might promote continued short-term improvement and long-term maintenance. For example, in the McDougall (2005a) study, it is possible that scheduling one phase during which performance was supposed to, and did, revert temporarily to a previously mastered level amounted to a well-timed physiological or psychological "break"—a behavioral buffer against satiation. Following this respite, the participant might have been well positioned to renew exercise efforts and perform the target behavior in accordance with criteria that became more stringent during subsequent intervention phases. Finally, like the CC, the RBCC is not conducive to interventions that are expected to produce huge, immediate changes in target behaviors (Hartmann & Hall, 1976; Poling & Fuqua, 1986; Schloss, Sedlak, Elliott, & Smothers, 1982).

The Distributed Criterion

The DC is the second recent design innovation based on the classic CC design. This design is particularly suited to empirical investigations of multitasking strategies, that is, where individuals allocate time to multiple interdependent tasks or contexts in ways that mesh with changing environmental demands. The DC incorporates features of the CC, multiple baseline, and ABAB designs. It typically includes a concurrent baseline phase across three contexts or behaviors, followed by a series of concurrent intervention phases, a priori performance criterion for the intervention phases (classic CC feature), multiple target behaviors or contexts (multiple-baseline feature), and reversal phases (ABAB feature). As with the CC, experimental control in the DC is demonstrated most clearly when the target behavior conforms quickly, precisely, and in a stable manner to changes in performance criteria across sequential (adjacent) intervention phases. However, experimental control in the DC also requires that interdependent target behaviors conform simultaneously to changes in performance criteria that are distributed across concurrent intervention phases. Thus, the DC is most useful for evaluating interventions that aim to change concurrently—through small- or large-level changes and in two directions—interdependent behavior across multiple contexts. The similarities of, and differences between, the DC and the classic CC design are provided in Table 2.

First Application of the Distributed Criterion

In the first study to apply the DC (McDougall, 2005c), the sole participant used a multitasking strategy that incorporated goal setting and behavioral self-management to increase research productivity from a few minutes per day during baseline to 4 hrs per day during the intervention phases. Research productivity was operationally defined and measured as the mean number of minutes (i.e., moving daily average) that the participant performed activities (e.g., data analysis, typing,

TABLE 2. Comparing and Contrasting the Changing Criterion and Distributed Criterion

Feature	Changing criterion	Distributed criterion
Number of target behaviors or contexts	$N = 1$, in one context	$N > 1$, or one target behavior is performed in > 1 context
Bolster case for experimental control via:	One phase that temporarily reverts or changes criterion in opposite-of-usual and non-therapeutic direction	One or more phases that change criterion in any direction, but without non- therapeutic threat
Design typically applied as:	Single design, sometimes with one phase that temporarily reverts performance criterion	Combined design with multiple baseline, changing criterion & ABAB elements
Number and duration of baseline phase(s)	One, usually brief	One per behavior/context, usually brief
How criteria are applied	One criterion at fixed value across all sessions within an intervention phase. Criterion shifts in step-wise manner for successive phases, for a single target behavior, in a single context, in one direction – either increases or decreases, but not in both directions. Change in criterion (and level change) is typically small.	Overall criterion is constant across all sessions across all, or nearly all, intervention phases. Criterion shifts can be large or small. Overall criterion is distributed across multiple individual behaviors or contexts for various intervention phases – in two directions, increases and decreases concurrently.
Amenable to interventions that:	Shape one behavior in one context in one desired direction; that is, to increase OR decrease one behavior. Utilize differential reinforcement of higher or lower rates of behavior, goal setting, behavioral self-management	Allocate or manage one behavior in more than one context, or multiple behaviors, in two desired directions – increase AND decrease; multi-task, prioritize, & re-allocate tasks based on demands, due dates, schedules, goal setting, behavioral self-management

Note. From "The Distributed Criterion Design," by D. McDougall, in press, *Journal of Behavioral Education*. Copyright 2005 by Springer-Verlag. Reprinted with permission.

editing) to complete three journal manuscripts. During baseline, the participant self-recorded research productivity for each of three manuscripts. During the intervention phases, the participant continued to self-record but also used goal setting and self-graphing. Goal setting required the participant to establish long-term and intermediate goals (e.g., dates by which manuscripts would be completed) and short-term objectives (STOs). STOs specified performance criteria, that is, how many minutes per day, on average, the participant would perform research activities (see Table 3). Self-graphing required the participant to post his research productivity on a line graph every day.

Distributing One Total Criterion Across Multiple Tasks. In McDougall (2005c), the overall criterion for the participant's total productivity within and across intervention phases was fixed at a mean of greater than or equal to 3 hrs per day, whereas productivity criteria for each of three individual manuscripts varied. That is, the overall criterion of at least 3 hrs was distributed across the three manuscripts in a manner consistent with the multitasking nature of the intervention and target behavior. Productivity criteria for individual manuscripts were shifted in accordance with the participant's need to devote varying amounts of time to manuscripts in various stages of development, ranging from initial drafts to nearly complete to complete. *Nearly complete* was operationally defined as the point at which the participant sent a manuscript to colleagues for feedback. At such a point, the participant reduced the productivity criterion for the nearly complete manuscript to 0 min/day. Concurrently, the participant modified productivity criteria, initiated work, and increased the time devoted to the other manuscripts. After receiving feedback from colleagues on a nearly complete manuscript, the participant renewed work on that manuscript, finished it, and submitted it to a journal for review. A *complete manuscript* was operationally defined as the point at which the participant mailed a manuscript to a journal for review.

Phase changes were instituted based on mastery of STOs that the participant established through goal setting. Labels for each intervention phase in Figure 2 correspond to each STO listed in Table 3. In this DC design variation, data collection included *interdependent* measures of productivity on three manuscripts. The DC design and interdependent measures were consistent with the multitasking strategy evaluated in this study and with the corresponding research question: To what extent does a multitasking strategy (i.e., goal setting combined with self-graphing) affect research productivity? McDougall (2005c) derived the multitasking strategy from re-

TABLE 3. Short-Term Objectives Established By Participant Correspond to Research Productivity Criteria for Successive Intervention Phases

Short-term objective	Intervention phase	Phase label	Performance expectations with mean number of minutes planned for research activities: Manuscript A	Manuscript B	Manuscript C
1	1	180A-0B-0C	Initiate work, 180/day	No work, 0/day	No work, 0/day
2	2	120A-60B-0C	Reduce work to 120/day until nearly complete	Initiate work, 60/day	No work, 0/day
3	3	0A-120B-60C	Suspend work, 0/day; await peer feedback	Increase work to 120/day until nearly complete	Initiate work, 60/day
4	4	120A-0B-60C	Renew work, 120/day until complete	Suspend work, 0/day; await peer feedback	Continue work, 60/day
5	5	XA-180B-0C	None = manuscript submitted to journal	Renew work, 180/day until complete	Suspend work, 0/day
6	6	XA-XB-180C	None = manuscript submitted to journal	None = manuscript submitted to journal	Renew work, 180/day until nearly complete
7	7	XA-XB-0C	None = manuscript submitted to journal	None = manuscript submitted to journal	Suspend work, 0/day await peer feedback
8	8	XA-XB-180C	None = manuscript submitted to journal	None = manuscript submitted to journal	Renew work, 180/day until complete

Note. The alpha-numeric sequences that appear in the "Phase Label" column include: (a) subscript letters (A, B, or C) that identify each of the three manuscripts, including Manuscript A, B, and C; and (b) corresponding numbers that specify the productivity criterion, in minutes, for each manuscript within various phases. The letter X indicates that a productivity criterion was no longer relevant because the participant had completed the manuscript and submitted it to a journal for review and publication. From "The Distributed Criterion Design," by D. McDougall, in press, *Journal of Behavioral Education*. Copyright 2005 by Springer-Verlag. Reprinted with permission.

search findings, including those of Boice (1990), who recommended that professors schedule and work concurrently on multiple writing projects because "alternatives reduce the tedium that can emerge when working on the same project too regularly" (p. 78).

Experimental Control and Phase Comparisons. A strength of the DC is that it provides numerous comparisons of both concurrent and sequential phases. These comparisons permit investigators to evaluate experimental control in studies that investigate multitasking with interdependent tasks—something that individual single-case designs (i.e., multiple baseline, ABAB, or CC) cannot accommodate. Sequentially derived comparisons, that is, changes in the target behavior across adjacent phases for individual manuscripts, are displayed in Figure 2. Four sequential comparisons are available for Manuscript A:

1. a comparison of productivity during baseline and the initial intervention phase (180-0-0);
2. an adjacent-phase comparison—initial intervention phase versus second intervention phase (120-60-0);
3. an adjacent-phase comparison—second intervention phase versus third intervention phase (0-120-60); and
4. an adjacent-phase comparison—third intervention phase versus fourth intervention phase (120-0-60).

Figure 2 reveals that for Manuscripts B and C, five and eight sequential comparisons, respectively, are available. Thus, a total of 17 sequentially derived comparisons are available for evaluating changes in research productivity across adjacent phases of the study.

In addition to sequential comparisons, the *distributed* component of the DC requires investigators to evaluate *concurrent* changes in interdependent performance. In Figure 2, examining the graphed data vertically across the three manuscripts shows changes in performance criteria that were implemented concurrently. For example, performance criteria (0-120-60) for the third intervention phase stipulated that the participant reduce productivity on Manuscript A to 0 min, concurrently increase productivity on Manuscript B to 120 min, and concurrently initiate productivity on Manuscript C at 60 min. The case for experimental control here depends largely on how the participant's behavior shifts in accordance with concurrent changes in performance criteria. In this example, the following questions should be asked:

1. To what extent does productivity conform to concurrently implemented changes in criteria

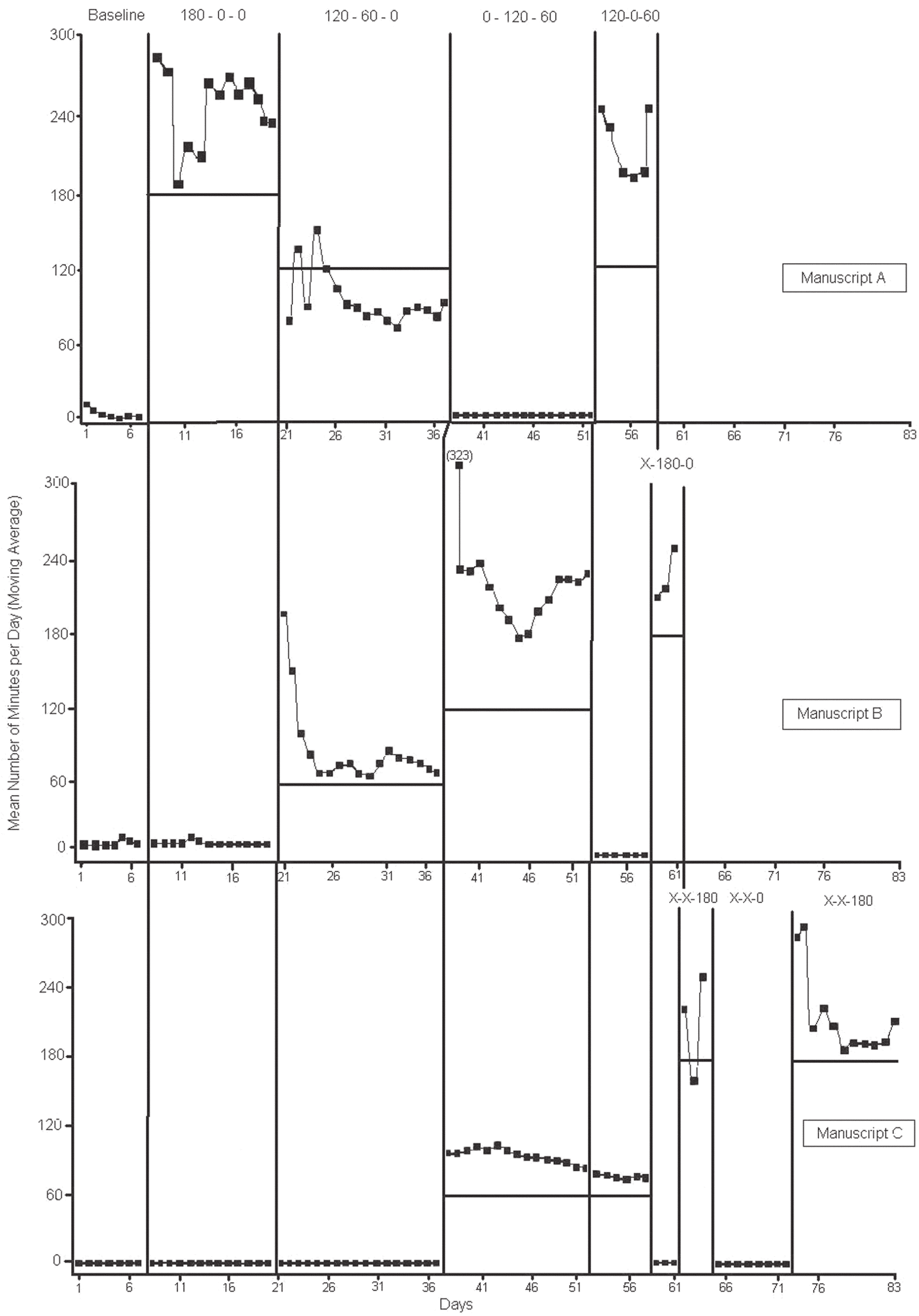

FIGURE 2. Moving average for research productivity (mean number of minutes expended daily) within baseline and intervention phases for Manuscripts A, B, and C. Note. Horizontal lines indicate within-phase productivity criteria (i.e., minimum number of minutes to be expended on manuscript) that participant set for self. Labels for intervention phases appear as numeric sequences (e.g., 120-60-0), with the first numeral indicating the within-phase criterion (mean number of minutes) that the participant established for Manuscript A, the second numeral indicating the within-phase criterion for Manuscript B, and the third numeral indicating the within-phase criterion for Manuscript C. X indicates that a criterion was no longer pertinent because the participant had completed work on that manuscript. From "The Distributed Criterion Design," by D. McDougall, in press, *Journal of Behavioral Education*. Copyright 2005 by Springer-Verlag. Reprinted with permission.

when the criterion for (a) the first manuscript decreases, (b) the second manuscript increases, and (c), the third target manuscript increases?

2. Does productivity conform to performance criteria for one, two, or three of the manuscripts? The case for experimental control is strongest when productivity on each of three manuscripts conforms to criterion shifts.

In this example (Phase 0-120-60 in Figure 2), concurrent changes in productivity provide some—but not definitive—support for experimental control. First, productivity for Manuscript A *decreased and matched* the 0-min/day criterion. Concurrently, productivity for Manuscript B *increased and exceeded* the 120-min/day criterion. At the same time, the participant initiated productivity for Manuscript C and *exceeded* the 60-min/day criterion. Experimental control would be stronger if productivity for Manuscript B (M = 231 min/day) exceeded the performance criterion of 120 min by only a few minutes rather than nearly doubling the criterion. Coincidentally, this example illustrates the potential benefits of using an RBCC design. Suppose this study incorporated the RBCC and defined a relatively narrow range (110–130 min) of expected productivity for Manuscript B instead of a single-point criterion (120 min). Productivity might have been more stable with the range-bound feature, and experimental control could have been stronger.

Because performance measures in DC designs are interdependent, researchers must evaluate data from concurrent phases. For example, the vertical panels in Figure 2 reveal that numerous concurrent phase comparisons are available. In this DC design, we must determine the number of manuscripts for which productivity conforms to concurrently implemented performance criteria for each of the first four intervention phases. We must also ascertain whether productivity for Manuscripts B and C conforms to concurrently implemented performance criteria for the fifth intervention phase (X-180-0).

Overall, visual inspection of graphed data and the numerous sequential and concurrent phase comparisons suggest that the multitasking intervention strategy demonstrated moderate to strong control over research productivity. As indicated in Figure 2, with one exception (mean productivity for Manuscript A during Phase 120-60-0 = 92 min), mean productivity matched or exceeded within-phase criteria (a) across adjacent phases for individual manuscripts and (b) across manuscripts when criterion changes for manuscripts were implemented concurrently. In addition, trends in Figure 2 indicate that the productivity criteria exerted control over the target behavior when descending trends and phase means approached or fell below performance criteria. In such cases, the participant (a) substantially increased productivity during subsequent days (see Phase 180-0-0 for Manuscript A, and Phase X-X-180 for Manuscript C) or (b) maintained productivity just enough to remain above criteria (see Phase 120-60-0 for Manuscript B, and Phase 120-0-60 for Manuscript C). As the graph also confirms, social validity data (i.e., the participant's subjective evaluation) indicated that the participant usually used a "get ahead and stay ahead" strategy. That is, mean productivity greatly exceeded criteria during the initial sessions of an intervention phase and then gravitated toward the within-phase performance criteria during the remaining sessions.

Strengths, Limitations, and Considerations

Perhaps the outstanding feature of the DC is the capacity to demonstrate experimental control for multitasking strategies and interdependent tasks. Three design elements contribute to this feature:

1. The DC provides numerous opportunities to replicate intervention impact.
2. The DC requires both sequential (i.e., across adjacent) phases and concurrent (i.e., across behaviors or contexts) changes in target behaviors as a function of intervention.
3. The DC requires bidirectional changes in target behaviors (i.e., both increases and decreases).

Another strength is that the design can accommodate concurrent implementation of initial intervention across multiple tasks, as in the example in Figure 2, or staggered implementation of initial intervention, as in a classic multiple-baseline design.

Unique Advantage of the DC Design

As illustrated in Figure 2, the DC accommodates bi-directional changes in performance criteria and temporary reversions of performance criteria to prior levels. These options offer investigators numerous opportunities to demonstrate experimental control. Notably, investigators can use these options without encountering most of the ethical, practical, and therapeutic concerns inherent in designs, such as the ABAB, that temporarily withdraw intervention, or some variations of the classic CC, which temporarily revert the behavior in a nontherapeutic direction. For example, an investigator might seek to reduce, in a stepwise fashion, the number of cigarettes smoked each day from 20 per day to 18 per day, and so forth. At some point, the investigator may revert temporarily to a higher criterion to establish a more convincing case for experimental control. If smoking behavior conforms to this temporary higher criterion, experimental control is strengthened, particularly if the behavior conforms to a series of stepwise reductions in the criterion during subsequent intervention phases. However, raising the criterion for number of cigarettes smoked presents ethical concerns and conflicts with the therapeutic goal of reducing smoking. Conversely, in a DC design, changes in the direction that is opposite from the usual for performance criteria (e.g., reducing the criteria for the

number of minutes a participant works on Manuscript A) are rarely, if ever, nontherapeutic. Why? Because the participant reallocates the time devoted previously to one task (e.g., Manuscript A) to work on other tasks (e.g., Manuscripts B and C). The overall criterion remains constant for multiple interdependent tasks (i.e., an overall moving average of 3 hrs per day for the three manuscripts combined).

Applying the DC in Special Education Research and Practice

General Case. We can envision many research questions and behavior change interventions for which the DC might be useful, such as interventions that aim to improve the performance of students who must execute multiple concurrent tasks, each of which might vary in importance, priority, or urgency across time. Furthermore, students often must perform these tasks under conditions that fix or limit the amount of time or resources they may allocate to the tasks.

Specific Examples. Possible applications of the DC for research in special education include interventions designed to (a) promote timely completion and management of academic assignments, (b) diversify leisure or recreational activities, (c) improve social interactions, and (d) improve acquisition and maintenance of basic academic skills, such as multiplication facts. Consider, for example, the multitasking challenges faced by high school students with and without disabilities. These students must learn how to prioritize, allocate, and reallocate their time and efforts to manage daily homework, intermediate-range tasks (e.g., an academic assignment due in 2 weeks), and long-term projects (e.g., an end-of-semester book report).

A second example involves leisure activities. Some students spend an inordinate amount of their time on one activity, such as playing video games or watching television. In such cases, diversifying leisure activities might be a useful goal. The DC could mesh with interventions designed to redistribute leisure time (i.e., reduce excessive time on one leisure activity while concurrently initiating and increasing time on two other leisure activities). In a third example—a student with social skills deficits and excesses—the DC could be applied to evaluate the effectiveness of a multitasking intervention designed to (a) decrease one social skill excess, such as frequent cursing, (b) establish a replacement behavior not currently in the student's repertoire (e.g., cursing inaudibly or vocalizing words that substitute for curse words), and (c) promote use of alternative behaviors that are in the student's repertoire but are used infrequently (e.g., expressing needs with "I" statements).

As a final example, investigators could use the DC to verify the efficacy of multitasking instructional and practice procedures designed to promote acquisition and maintenance of basic multiplication facts. Given a fixed amount of daily practice time, teachers or students might begin by spending all of it mastering the easiest fact sets (0x, 1x). As these easy fact sets are mastered, practice time can be redistributed; that is, students would reduce their practice time to a maintenance level on the easiest (i.e., mastered) sets while concurrently initiating practice time on subsequent sets in the series (e.g., 5x, 2x).

Summary of DC Design

The DC incorporates elements of the CC, multiple baseline, and ABAB designs. It permits investigators to evaluate experimental control via numerous replications. One advantage of the DC is that it allows researchers to investigate behavioral clusters and questions of interest in multitasking contexts—real-world contexts that confront students with disabilities on a daily basis. Like all students, students with disabilities must learn how to juggle multiple, concurrent, and overlapping tasks. Because they often must perform such tasks under conditions that limit or fix the amount of time and resources they can allocate to tasks, they must learn to prioritize, allocate, adjust, and readjust their efforts to maximize efficiency, particularly when circumstances and task demands change, abate, or emerge. Investigators can use the DC to evaluate experimental control in studies that target interdependent performance in ever-changing, multitasking contexts. Moreover, teachers and students can apply the DC approach as a scheduling and time-management tool.

By virtue of its design requirements, the DC is more complex and will have fewer applications compared to basic, single-case research designs, especially the versatile multiple-baseline design. To maximize experimental control and minimize threats to internal validity, researchers must adapt guidelines from the CC, multiple baseline, and ABAB designs based on nuances that arise when they integrate elements of these three designs into the DC.

Conclusion

Clinical needs influenced the design features of the two recent variations of the classic CC design described in this article. In the RBCC, a range-bound feature was incorporated initially as a clinical strategy (McDougall, 2005a). This feature (a) permitted some flexibility in duration of daily exercise, (b) restricted the duration of daily exercise to a healthy range, (c) helped to control how much the participant increased his exercise, and (d) minimized the likelihood of injury associated with exercising too much, too soon, and too variably. In the DC, clinical conditions required the participant to distribute work in a multitasking context: The participant worked concurrently and in overlapping fashion, in multiple contexts, and under conditions that limited and fixed total allocable time (McDougall, in press). By incorporating features of the classic CC, multiple baseline, and ABAB designs, the DC provided

a more responsive and conducive approach to investigating multitasking when compared to each design individually. As such, the DC bolsters Osborne and Himadi's (1990) claim that the CC design "possesses the flexibility to include other single-subject design elements" (p. 81). The DC can also accommodate the complexity of factors that affect the internal validity of combined single-case interventions.

We hope this article stimulates researchers and practitioners' awareness, understanding, and use of the RBCC and DC in special education. To date, neither innovation has appeared in classic (cf. Kazdin, 1982) or recent (cf. Kennedy, 2005) texts on single-case research designs, or in the professional literature in special education. Given that the RBCC is a very simple variation of Sidman's (1960) classic CC design, we are surprised that it has appeared only recently in a professional journal (McDougall, 2005b). We also hope this article promotes greater overall use of the classic CC design, which may be underutilized (Osborne & Himadi, 1990). Indeed, research-validated instructional practices and behavioral principles (e.g., setting explicit standards and shaping terminal behaviors through successive approximations) underscore CC designs. Finally, we recommend that researchers and practitioners consult sources on single-case research, as well as strategies, particularly goal setting, shaping, and self-management, that mesh well with CC design variations (cf. Alberto & Troutman, 1999; Barlow & Hersen, 1984; Glynn, Thomas, & Shee, 1973; Johnston & Pennypacker, 1993; Kazdin, 1982; Kennedy, 2005; Kratochwill, 1978; McDougall, 1998; McDougall, Smith, Black, & Rumrill, in 2005; Poling & Fuqua, 1986; Tawney & Gast, 1984; Watson & Tharp, 2002).

REFERENCES

Alberto, P. A., & Troutman, A. C. (1999). *Applied behavior analysis for teachers* (5th ed.). Columbus, OH: Merrill.

Baer, D. M., Wolf, M. M., & Risley, T. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*, 91–97.

Barlow, D. H., & Hersen, M. (1984). *Single-case experimental design: Strategies for studying behavior change* (2nd ed.). Elmsford, NY: Pergamon Press.

Bates, P., Renzaglia, A., & Clees, T. J. (1980). Improving the work performance of severely/profoundly retarded young adults: The use of a changing criterion procedural design. *Education and Training of the Mentally Retarded*, *15*(2), 95–104.

Boice, R. (1990). *Professors as writers: A self-help guide to productive writing.* Stillwater, OK: New Forums Press.

Edinger, J. D. (1978). Modification of smoking behavior in a correctional institution. *Journal of Clinical Psychology*, *34*(4), 991–998.

Flood, W. A., & Wilder, D. A. (2004). The use of differential reinforcement and fading to increase time away from a caregiver in a child with separation anxiety disorder. *Education & Treatment of Children*, *27*(1), 1–8.

Foxx, R. M., & Rubinoff, A. (1979). Behavioral treatment of caffeinism: Reducing excessive coffee drinking. *Journal of Applied Behavior Analysis*, *12*, 335–344.

Glynn, E. J., Thomas, J. D., & Shee, S. M. (1973). Behavioral self-control of on-task behavior in an elementary classroom. *Journal of Applied Behavior Analysis*, *6*, 105–113.

Gorski, J. A. B., & Westbrook, A. C. (2002). Use of differential reinforcement to treat medical non-compliance in a pediatric patient with leukocyte adhesion deficiency. *Pediatric Rehabilitation*, *5*(1), 29–35.

Hall, R. V. (1971). *Managing behavior: Behavior modification, the measurement of behavior.* Lawrence, KS: H & H Enterprises.

Hall, R. V., & Foxx, R. G. (1977). Changing criterion designs: An alternative applied behavior analysis procedure. In C. C. Etzel, G. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, mind, and application* (pp. 151–166). Hillsdale, NJ: Erlbaum.

Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, *9*, 527–532.

Hinerman, P. S., Jenson, W. R., Walker, G. R., & Petersen, B. P. (1982). Positive practice overcorrection combined with additional procedures to teach signed words to an autistic child. *Journal of Autism and Developmental Disorders*, *12*(3), 253–263.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Kahng, S., Boscoe, J. H., & Byrne, S. (2003). The use of an escape contingency and a token reinforcement economy to increase food acceptance. *Journal of Applied Behavior Analysis*, *36*, 349–353.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston: Allyn & Bacon.

Kerlinger, F. N. (1986). *Foundations of behavioral science*. New York: CBS College Publishing.

Kottler, J. A. (2001). *Making changes.* Philadelphia: Brunner-Routledge.

Kratochwill, T. R. (1978). *Single subject research.* New York: Academic Press.

Luiselli, J. K. (2000). Cueing, demand fading, and positive reinforcement to establish self-feeding and oral consumption in a child with chronic food refusal. *Behavior Modification*, *24*(3), 348–358.

McDougall, D. (1998). Research on self-management techniques used by students with disabilities in general education settings: A descriptive review. *Remedial and Special Education*, *19*, 310–320.

McDougall, D. (2005a) *The impact of goal setting and behavioral self-management on daily exercise and health of an overweight, middle-age adult.* Manuscript submitted for publication.

McDougall, D. (2005b). The range-bound changing criterion design. *Behavioral Interventions*, *20*, 129–137.

McDougall, D. (2005c). *Using goal setting and behavioral self-management to increase research productivity.* Manuscript submitted for publication.

McDougall, D. (in press).The distributed criterion design. *Journal of Behavioral Education.*

McDougall, D.,Smith, G., Black, R., & rumrill, P. (2005). Recent innovations in small-N designs for rehabilitation research: An extension of Cowan, Hennessey, Vierstra, and Rumrill. *Journal of Vocational Rehabilitation,* 197-205.

McLaughlin, T. F. (1983). An examination and evaluation of single subject designs used in behavior analysis research in school settings. *Educational Research Quarterly*, *7*, 35–42.

Mizes, J. S. (1985). The use of contingent reinforcement in the treatment of a conversion disorder: A multiple baseline study. *Journal of Behavior Therapy and Experimental Psychiatry*, *16*, 341–345.

Noles, K. O., Epstein, L. H., & Jones, H. T. (1976). Instructional control of sleep reduction. *Perceptual and Motor Skills*, *43*, 669–670.

Osborne, M. L., & Himadi, B. (1990). Evaluation of a shaping procedure with the changing criterion design. *Behavioral Residential Treatment*, *5*(2), 75–81.

Paniagua, F. A., Pumariega, A. J., & Black, S. A. (1988). Clinical effects of correspondence training in the management of hyperactive children. *Behavioral Residential Treatment*, *3*(1), 19–40.

Poling, A., & Fuqua, R. W. (1986). *Research methods in applied behavior analysis: Issues and advances.* New York: Plenum Press.

Schleien, S. J., Wehman, P., & Kiernan, J. (1981). Teaching adult leisure skills

to severely handicapped adults: An age-appropriate darts game. *Journal of Applied Behavior Analysis*, *14*, 513–519.

Schloss, P. J., Sedlak, R. A., Elliott, C., & Smothers, M. (1982). Application of the changing- criterion design in special education. *The Journal of Special Education*, *16*, 359–367.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Skinner, C. H., Skinner, A. L., & Armstrong, K. J. (2000). Analysis of a client–staff-developed shaping program designed to enhance reading persistence in an adult diagnosed with schizophrenia. *Psychiatric Rehabilitation Journal*, *24*(1), 52–57.

Tawney, J., & Gast, D. (1984). *Single subject research in special education*. Columbus, OH: Merrill.

Watson, D. L., & Tharp, R. G. (2002). *Self-directed behavior: Self-modification for personal development*. Belmont, CA: Wadsworth.

Weis, L., & Hall, R. V. (1971). Modification of cigarette smoking through avoidance of punishment. In R. V. Hall (Ed.), *Managing behavior: Behavior modification applications in school and home* (pp. 77–102). Lawrence, KS: H & H Enterprises.